

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-039325

(43)Date of publication of application : 12.02.1999

(51)Int.Cl. G06F 17/30
G06F 15/18
G06T 7/00
G10L 7/04
H04N 5/76

(21)Application number : 09-195356 (71)Applicant : MATSUSHITA ELECTRIC IND
CO LTD

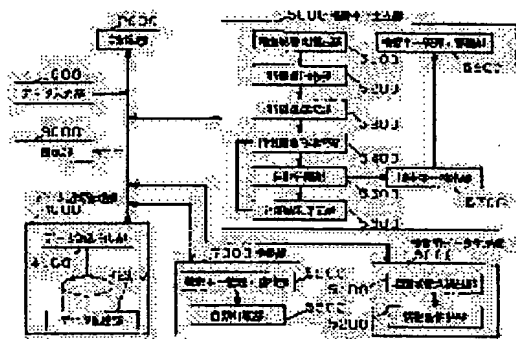
(22)Date of filing : 22.07.1997 (72)Inventor : AKIMOTO TOSHIKI
OKA NATSUKI

(54) SIMILARITY RETRIEVAL METHOD AND SYSTEM THEREFOR

(57)Abstract:

PROBLEM TO BE SOLVED: To extract the result obtained by learning so that a sample picture becomes the same category where retrieval intention is reflected as a retrieval key and to improve retrieval efficiency.

SOLUTION: This system is provided with a feature quantity extraction means 5200 for extracting the feature quantity of plural retrieval key candidate videos containing more than two designated retrieval key candidate videos, a feature quantity weight learning means 5400 for learning the weight of extracted feature quantity, an automatic classification means 5500 for classifying weighted feature quantity by using a neural circuit network model, a classified result evaluation means 5600 for evaluating the classified result and a retrieval key extraction means 5700 for extracting feature quantity representing the category, the weight and the weight of the link of an input/output unit as the retrieval keys. The video constituted of the similar picture and sound is retrieved by the retrieval key generated by learning so that the retrieval key candidate video designated among the retrieval key candidate videos is similarly classified.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

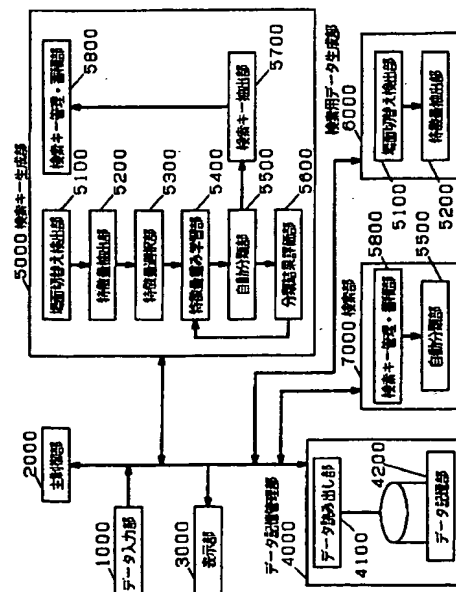
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

(11)特許出願公開番号



【特許請求の範囲】

【請求項1】 検索キーにより類似した映像を検索する類似検索方法において、検索キー候補映像の中から指定した検索キー候補映像が同一分類となるように学習により検索キーを生成することを特徴とする類似検索方法。

【請求項2】 検索キーは、指定した2つ以上の検索キー候補映像を含む複数の検索キー候補映像の特徴量に対して、指定された検索キー候補映像同士と指定された検索キー候補映像とそれ以外の検索キー候補映像の組合せで特徴量の重み学習と学習した特徴量の重み付けを用いて分類および分類結果の評価を繰り返し、指定した検索キー候補映像が同一分類となるように学習した結果を検索キーとして生成することを特徴とする請求項1記載の類似検索方法。

【請求項3】 学習した特徴量の重み付けを用いた分類は、神経回路網モデルを用いたことを特徴とする請求項2記載の類似検索方法。

【請求項4】 特徴量は、画像の特徴量と音の特徴量を組み合わせたことを特徴とする請求項1または2記載の類似検索方法。

【請求項5】 画像の特徴量は、色のヒストグラム、エッジ画素情報、2つのフレームのエッジ画素変化情報の一つまたは複数の組み合わせることを特徴とする請求項4記載の類似検索方法。

【請求項6】 音の特徴量は、可聴音を32のサブバンドに分割し、サブバンド毎に、場面全体の平均音量、最大音量、音量分散値、または周波数成分、ケフレンシー（周波数成分を対数変換した値を逆周波数変換した成分）、音量を一つまたは複数の組み合わせることを特徴とする請求項4記載の類似検索方法。

【請求項7】 画像または音からの検索意図を特徴づける特徴量は、映像の場面の変わり目のフレームから求めることを特徴とする請求項2乃至6のいずれかに記載の類似検索方法。

【請求項8】 検索キーは、同一分類となるように学習した結果として、カテゴリを代表する特徴量とその重み並びに入出力ユニットのリンク重みであることを特徴とする請求項2乃至7のいずれかに記載の類似検索方法。

【請求項9】 特徴量は、画像または音からの検索意図を特徴づける特徴量から少なくとも1つ以上の特徴量を選択することを特徴とする請求項4乃至6のいずれかに記載の類似検索方法。

【請求項10】 検索キーにより類似した画像または音を検索する類似検索方法において、検索対象映像は予め特徴量を抽出した検索用データとして求めておくことを特徴とする請求項1または2記載の類似検索方法。

【請求項11】 検索キーには、検索キー候補映像の場面を反映したタグを付けることを特徴とする請求項1または10記載の類似検索方法。

【請求項12】 指定した2つ以上の検索キー候補映像

を含む複数の検索キー候補映像の特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量に対して、指定された検索キー候補映像同士と指定された検索キー候補映像とそれ以外の検索キー候補映像の組合せで特徴量の重みを学習する特徴量重み学習手段と、前記特徴量重み学習手段で重みづけされた特徴量を神経回路網モデルを用いて分類する自動分類手段と、前記自動分類手段で分類した結果を評価する分類結果評価手段と、前記特徴量重み学習手段からのカテゴリを代表する特徴量とその重みと前記自動分類手段からのリンクの重みを検索キーとして抽出する検索キー抽出手段とを備え、検索キー候補映像の中から指定した検索キー候補映像が同一分類となるように学習により生成した検索キーにより類似した画像と音からなる映像を検索することを特徴とする類似検索システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、映像情報を記憶している媒体から、利用者が求める場面に類似したサンプル映像から学習により生成した検索キーにより映像情報を検索する類似検索方法及び類似検索システムに関するものである。

【0002】

【従来の技術】近年、オーサリングシステムのように、テキスト、静止画、サウンド、AVデータ等の様々なデータを扱うシステムにおいて、利用者が望むデータを効率良く獲得できる検索システムが望まれている。現在、キーワードを組み合わせた明示的な検索方法以外に、あいまいな検索意図を検索に適用するための方法がいくつか提案されている。類似検索は、その一例であり、自分の検索意図に適合したサンプルを検索キーとして、検索キーに類似したデータを検索対象とする方法である。従来、類似検索システムとしては、特開平9-44518号公報に開示されたものが知られている。

【0003】図8は、従来の類似検索システムのブロック構成図であり、図9は従来の類似検索システムの処理を説明するフローチャートである。図8において、画像データベースに画像データを登録する画像データベース登録モジュール4と、上記画像データベース1から検索条件に合う画像データを検索する画像データベース検索モジュール5と、ファジィルールを適用するファジィ推論手段3とからなる。上記画像データベース登録モジュール4は、画像データから画像の特徴ベクトルを抽出する特徴ベクトル抽出手段41と、上記特徴ベクトルを入力し、ファジィルールを適用して高次の特徴ベクトルを生成する高次特徴ベクトル生成手段42と、上記画像データの上記特徴ベクトルを入力し、ファジィルールを適用して上記入力された特徴ベクトルをパターンに分類するパターン分類手段43とからなる。上記画像データベース検索モジュール5は、上記画像データベース1に登

録された上記画像データの縮小アブストラクト画像を表示し、上記表示された画像の中から検索用画像を選択し、上記選択された画像の画像データに基づいて検索ようデータを生成する検索用データ生成手段51と、上記検索用生成手段51により生成された上記検索用データを入力し、上記ファジールールを適用して上記画像データベース1に登録された画像データを検索するファジ検索手段52とからなる。

【0004】上記のように構成された従来の類似検索システムの動作を図9のフローチャートを用いて以下に説明する。

【0005】検索用の画像から上記画像の特徴量を表わす第1の特徴ベクトルを抽出する特徴抽出処理（ステップ10）と、上記画像の抽出された第1の特徴ベクトルをファジールールに従って階層的なカテゴリにパターン分類するパターン分類処理（ステップ20）と、上記第1の特徴ベクトルと上記分類されたパターンから登録用データを生成し、画像データベースに登録する画像データ登録処理（ステップ30）とからなる。順次に高次の特徴ベクトルを抽出して、ステップ10からステップ30を繰り返し行なうことにより、階層的なパターン分類が行われる。

【0006】更に、ファジールールを用いて階層的なカテゴリにパターン分類された画像の特徴ベクトルに応じて、上記画像データベースに分類して登録された上記画像データの縮小されたアブストラクト画像を表示する画像表示処理（ステップ40）と、上記表示された画像の中から検索対象画像と類似した少なくとも一つの画像を選択する検索条件指定処理（ステップ50）と、上記選択された少なくとも一つの画像の上記特徴ベクトルから検索用データを生成する検索用データ生成処理（ステップ60）と、上記生成された検索用データを検索条件として上記画像データベースをファジールールに基づいて検索し、上記検索用データに対応する上記画像データが分類されたパターンを得るファジ検索処理（ステップ70）とからなる。

【0007】このように、ファジールールを用いて検索キーを階層的にパターン分類し、ファジ検索することによって、あいまいな検索を行なうことができる。

【0008】

【発明が解決しようとする課題】しかし、パターン分類された結果は、特徴量とパターン分類方法により決定されるものであり、必ずしも検索意図と一致した分類にはならない。また、静止画や映像は、見方（検索意図）を変えることによって複数のカテゴリに属することができるが、階層的な分類方法では、多様な見方に十分対応できない。つまり、検索キーに類似したデータは、カテゴリの多様性を反映して一意に決めることはできないので、従来の方法では効率良く検索することができない。

【0009】この類似検索システムにおいては、検索意

図を反映したカテゴリに相当する検索キーを抽出することが求められている。

【0010】本発明は、検索意図を特徴づける特徴量の重み学習と、神経回路網モデルを用いた分類及び分類結果の評価を繰り返すことにより、サンプル画像が検索意図を反映した同一のカテゴリとなるように学習で得られた結果を検索キーとして抽出し、検索効率を向上できることを目的とする。

【0011】

10 【課題を解決するための手段】この課題を解決するために本発明は、検索キーにより類似した映像を検索する類似検索方法において、検索キー候補映像の中から指定した検索キー候補映像が同一分類となるように学習により検索キーを生成するものである。

【0012】また、指定した2つ以上の検索キー候補映像を含む複数の検索キー候補映像の特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量に対して、指定された検索キー候補映像同士と指定された検索キー候補映像とそれ以外の検索キー候補映像の組合せで特徴量の重みを学習する特徴量重み学習手段と、前記特徴量重み学習手段で重みづけされた特徴量を神経回路網モデルを用いて分類する自動分類手段と、前記自動分類手段で分類した結果を評価する分類結果評価手段と、前記特徴量重み学習手段からのカテゴリを代表する特徴量とその重みと前記自動分類手段からのリンクの重みを検索キーとして抽出する検索キー抽出手段とを備え、検索キー候補映像の中から指定した検索キー候補映像が同一分類となるように学習により生成した検索キーにより類似した画像と音からなる映像を検索するものである。

30 【0013】本発明によれば、検索意図を特徴づける特徴量の重み学習と、神経回路網モデルを用いた分類及び分類結果の評価を繰り返すことにより、サンプル画像が検索意図を反映した同一カテゴリとなるように学習で得られた検索キーにより、類似した画像と音からなる映像を検索するもので、検索効率を向上できる。

【0014】

40 【発明の実施の形態】本発明の請求項1に記載の発明は、検索キーにより類似した画像または音を検索する類似検索方法において、検索キー候補の中から検索したい画像または音に類似した検索キーを生成することを特徴とするものであり、検索意図を反映したカテゴリの特徴量を検索キーとして抽出するという作用を有する。

【0015】請求項2に記載の発明は、検索キーは、指定した2つ以上の検索キー候補映像を含む複数の検索キー候補映像の特徴量に対して、指定された検索キー候補映像同士と指定された検索キー候補映像とそれ以外の検索キー候補映像の組合せで特徴量の重み学習と学習した特徴量の重み付けを用いて分類および分類結果の評価を繰り返し、指定した検索キー候補映像が同一分類となる

ように学習した結果を検索キーとして生成することを特徴とするものであり、検索意図を特徴づける特徴量の重み学習と分類及び分類結果の評価を繰り返すことにより、検索意図を反映したカテゴリの特徴量を検索キーとして抽出するという作用を有する。

【0016】請求項3に記載の発明は、学習した特徴量の重み付けを用いた分類は、神経回路網モデルを用いたことを特徴とするもので、学習した結果を検索キーとして利用できるという作用を有する。

【0017】請求項4に記載の発明は、特徴量は、画像の特徴量と音の特徴量を組み合わせたことを特徴とするもので、画像と音の特徴量を用いることにより、多様な分類を行なうという作用を有する。

【0018】請求項5に記載の発明は、画像の特徴量は、色のヒストグラム、エッジ画素情報、2つのフレームのエッジ画素変化情報の一つまたは複数を組み合わせることを特徴とするもので、映像による多様な分類を行なうという作用を有する。

【0019】請求項6に記載の発明は、音の特徴量は、可聴音を32のサブバンドに分割し、サブバンド毎に、場面全体の平均音量、最大音量、音量分散値、または周波数成分、ケフレンシー（周波数成分を対数変換した値を逆周波数変換した成分）、音量を一つまたは複数を組み合わせることを特徴とするもので、音による多様な分類を行なうという作用を有する。

【0020】請求項7に記載の発明は、画像または音からの検索意図を特徴づける特徴量は、映像の場面の変わり目のフレームから求めることを特徴とするもので、動画像への適用を可能にするという作用を有する。

【0021】請求項8に記載の発明は、検索キーは、カテゴリを代表する特徴量とその重み並びに入出力ユニットのリンクの重みであることを特徴とするもので、学習した結果が検索意図を反映したものであり、学習した結果を検索キーとして利用できるという作用を有する。

【0022】請求項9に記載の発明は、特徴量は、画像または音からの検索意図を特徴づける特徴量から少なくとも1つ以上の特徴量を選択することを特徴とするもので、特徴量の組み合わせを自由に換えられるという作用を有する。

【0023】請求項10に記載の発明は、検索キーにより類似した画像または音を検索する類似検索方法において、検索対象映像は予め特徴量を抽出した検索用データとして求めておくことを特徴とするもので、予め検索用データとして変換しておくことにより処理時間の大幅な短縮を図ることができるという作用を有する。

【0024】請求項11に記載の発明は、検索キーには、検索キー候補映像の場面を反映したタグを付けることを特徴とするもので、場面を反映したタグにより容易に検索キーを指定することができる。

【0025】請求項12に記載の発明は、指定した2つ

以上の検索キー候補映像を含む複数の検索キー候補映像の特徴量を抽出する特徴量抽出手段と、前記特徴量抽出手段で抽出された特徴量に対して、指定された検索キー候補映像同士と指定された検索キー候補映像とそれ以外の検索キー候補映像の組合せで特徴量の重みを学習する特徴量重み学習手段と、前記特徴量重み学習手段で重みづけされた特徴量を神経回路網モデルを用いて分類する自動分類手段と、前記自動分類手段で分類した結果を評価する分類結果評価手段と、前記特徴量重み学習手段からのカテゴリを代表する特徴量とその重みと前記自動分類手段からのリンクの重みを検索キーとして抽出する検索キー抽出手段とを備え、検索キー候補映像の中から指定した検索キー候補映像が同一分類となるように学習により生成した検索キーにより類似した画像と音からなる映像を検索するもので、検索意図を特徴づける特徴量の重み学習と、神経回路網を用いた分類及び分類結果の評価を繰り返すことにより、検索意図を反映したカテゴリの特徴量を検索キーとして抽出するという作用を有する。

【0026】以下、本発明の実施の形態について、図を用いて説明する。

（実施の形態）図1は、本発明の実施の形態の類似検索システムの構成図を示す。図1において、5000は検索キー候補映像の中から利用者が指定した複数の検索キー候補映像が同一分類となるように学習した結果を検索キーとして生成する検索キー生成部、4000は検索対象となる映像、検索キー候補映像及び教師データを記憶、管理するデータ記憶管理部、3000は検索キー候補映像や検索の結果見つかった映像や検索キー生成部5000で抽出された検索キーの表示を行なう表示部、1000はキーボードやファイルなどからの、映像の検索指示や、検索キー生成部5000で抽出した検索キーに対応する場面を表示する指示や、検索キー生成部5000で行なう学習に必要なパラメータの入力や検索動作の指示を行なうデータ入力部、6000は検索対象となる映像はデータ記憶管理部4000から読み出され映像から特徴的なデータを抽出する検索用データ生成部、7000はデータ記憶管理部4000から読み出した検出用データを検索キー生成部5000で生成した検索キーを用いて類似検索を行う検索部、2000はシステムの全ての制御を行なう主制御部とから構成する。

【0027】また、検索キー生成部5000は、指定した検索キー候補映像を含む全ての検索キー候補映像の画像の色やエッジ強度の変化を分析することによって場面の変わり目を検出する場面切替え検出部5100と、場面切替え検出部5100によって検出された場面単位に、場面的特徴量を抽出する特徴量抽出部5200と、特徴量抽出部5200で抽出された特徴量の中から、データ入力部1000を使って指示された特徴量を選択する特徴量選択部5300と、特徴量選択部5300で選択された特徴量の重みを分類結果の評価を用いて特徴量の重みを学習する特徴量重み学習部54

00と、特徴量重み学習部5400で重みづけされた特徴量を、神経回路網モデルの1つである自己組織化マップと呼ばれるニューラルネットワークを用いてリンク重みを学習し、分類する自動分類部5500と、自動分類部5500で分類した結果を合致検索キー候補（利用者が指定した複数の検索キー候補映像）が含まれるカテゴリにおいて合致検索キー候補が含まれる割合を計算することにより評価する分類結果評価部5600と、自動分類部5500で分類した結果から代表的な特徴量を抽出し、その代表的な特徴量と特徴量重みさらに自動分類部5500で学習したリンク重みを検索キーとして抽出する検索キー抽出部5700と、検索キー抽出部5700で抽出された検索キーを蓄積・管理する検索キー管理・蓄積部5800とを備える。

【0028】検索用データ生成部6000は、映像の色やエッジ強度の変化を分析することによって場面の変わり目を検出する場面切替え検出部5100と、場面切替え検出部5100によって検出された場面単位に、場面的特徴量を抽出する特徴量抽出部5200とを備える。

【0029】検索部7000は、検索キー生成部5000で抽出された検索キーを蓄積・管理する検索キー管理・蓄積部5800と、データ記憶管理部4000から読み出した検索用データを検索キーの重み付けされた特徴量を用いて神経回路網モデルの1つである自己組織化マップと呼ばれるニューラルネットワークを用いて分類する自動分類部5500とを備える。

【0030】以上のように構成された類似検索システムについて、以下にその動作を述べる。

【0031】類似検索システムの動作は、検索キー抽出動作、検索用データ生成動作、検索動作の3つに分類される。

【0032】まず、検索キー生成部5000の検索キー抽出動作について説明する。検索キー生成部5000の検索キー抽出動作は、検索動作前に行われ、既に検索キー抽出動作が完了している場合には、データ入力部1000からの指示により、抽出が完了した検索キーの指定を検索キー生成部5000に対し行ない、検索キー抽出動作は行なわない。

【0033】検索キー抽出動作は、入力部1000から検索キー抽出の指示が入力されることによって開始する。主制御部2000は、検索キー生成部5000に対して、検索キー抽出命令を出し、続いてデータ記憶管理部4000に対し、検索キー候補映像の読み出しを指示し、読み出した検索キー候補映像の表示を表示部3000に指示する。表示部3000に表示された検索キー候補映像を見て利用者は、データ入力部1000を用いて、検索したい映像に合致する検索キー候補映像（以下合致検索キー候補）を2つ以上指定する。

【0034】これを受けて検索キー生成部5000は、検索キー抽出を行うもので、以下に詳細に説明する。

【0035】場面検出部5100は、合致検索キー候補を含

む検索キー候補映像の場面切替え検出動作を行なうもので、場面切替え検出結果を場面的先頭フレーム番号あるいは場面的先頭の時間として特徴量抽出部5200に出力するものである。場面検出部5100の場面切替え検出動作について、図2のフローチャートを用いて説明する。

【0036】まず、データ記憶管理部4000から合致検索キー候補を含む検索キー候補映像の時系列のフレーム画像を逐次読み出す（ステップ1）。読み出しの終了かどうかを判断し、終了の場合は動作を終了する（ステップ2）。

【0037】場面切替え（次シーン）候補及び、場面切替え型を検出する。場面切替え型には、4種類あり、場面切り替わりの時間の短い短時間長型とそれ以外の型に分類され、さらに、短時間長型以外のゆっくり場面的切り替わる型には、前シーンまたは次シーンが拡大・縮小・変形・移動する映像移動型、前シーンと次シーンの間の対応する画素を合成しながら切り替わる画素合成型、前シーンの映像の一部分が次シーンの対応する画素に置換され、置換される画素を拡大しながら切り替わる画素置換型がある（ステップ3）。場面切替え（次シーン）候補があるかどうかを判断し、ない場合はステップ1に戻る（ステップ4）。

【0038】場面切替え（次シーン）候補が検出された場合に、場面切替え（次シーン）候補の先頭位置の検出を行い、先頭位置として適当な条件を満たさなかった場合には、場面切替えとはみなされない（ステップ5）。先頭位置が検出されたかどうかを判断し、検出されない場合はステップ1に戻る（ステップ6）。

【0039】先頭位置を検出した場合には、先頭フレーム画像と前シーンの先頭画像の例えば色の類似度を求め、類似度の値がしきい値以下の場合、場面切替えと判断する（ステップ7）。

【0040】先頭位置は、場面的切替え型によって異なり、短時間長型は、色の類似度の時間変化を、また、映像移動型、画素合成型は輝度変化量の大きい画素の時間変化を、画素合成型は画面全体のエッジ強度の時間変化を用いて検出し、検出されなかった場合はステップ1に戻る（ステップ8）。先頭位置が検出された場合は、先頭のフレーム番号を通知しステップ1に戻る（ステップ9）。

【0041】なお、場面切替え検出方法に関しては、文献「編集効果を含む映像のシーンチェンジ検出方法、マルチメディアと映像処理シンポジウム'94（テレビジョン学会）、p.21-26(1994)」に詳しい。

【0042】なお、場面切替え処理については、さまざまな方式が提案されているが、本発明において、場面切替え処理方法は上記場面切替え処理に限定するものではない。

【0043】次に、特徴抽出処理部5200の特徴量抽出動作について詳細に説明する。特徴抽出処理部5200は、場

10

20

30

40

50

面単位に映像や音の情報を分析処理し、特徴量選択部5300に出力するものである。特徴抽出処理部5200の特徴量抽出動作は、場面切替えの単位に、映像、音に関して *

＊(表1)に示すような特徴量を抽出する。
【0044】
【表1】

メディア種別	特徴量種別	説明
映像	特徴量1	場面先頭フレームの色ヒストグラム
映像	特徴量2	場面先頭フレームのエッジ画素情報
映像	特徴量3	エッジ画素変化情報 (場面先頭フレームと10フレーム目)
映像	特徴量4	エッジ画素変化情報 (場面末尾10フレーム目と20フレーム目)
映像	特徴量5	エッジ画素変化情報 (場面末尾10フレーム前と末尾フレーム)
音	特徴量6	サブバンド毎の音量平均
音	特徴量7	サブバンド毎の音量最大値
音	特徴量8	サブバンド毎の音量分散
音	特徴量9	ケフレンシー分布パターン(最小時間成分最大時)
音	特徴量10	ケフレンシー分布パターン平均
音	特徴量11	場面先頭200msec分の音量平均
音	特徴量12	場面末尾200msec分の音量平均
音	特徴量13	場面先頭500msecから1000msecまでの音量平均
音	特徴量14	場面末尾1000msec前から500msec前までの音量平均
音	特徴量15	高域音持続時間

【0045】映像については、場面の先頭フレームの色ヒストグラム(特徴量1)、場面先頭フレームのエッジ画素情報(特徴量2)、先頭フレームと先頭から10フレーム目のエッジ画素変化情報(特徴量3)、先頭から10フレーム目と先頭から20フレーム目のエッジ画素変化情報(特徴量4)、場面末尾10フレーム前と末尾フレームのエッジ画素変化情報(特徴量5)を抽出する。エッジ画素変化情報とは、フレーム上の同じ位置にある画素に対し、変化前の画素がエッジであるかどうか(1ビット)、変化後の画素がエッジであるかどうか(1ビット)を表す情報であり、1画素あたり合計2ビットで表現できる。

【0046】なお、フレームを分割し、分割画面単位で上記特徴量1から特徴量5までを抽出しても実施可能である。また、以上の特徴量1及び特徴量2の説明では、場面の先頭フレームを用いて説明したが、任意のフレームについても同様に実施可能である。特徴量3、特徴量4及び特徴量5の説明で用いたフレームの組み合わせ以外でも実施可能である。

【0047】音については、20kHzまでの可聴音を32のサブバンドに分割し、サブバンド毎に、場面全体の音量平均(特徴量6)、音量最大値(特徴量7)、音量分散(特徴量8)、500msecの切り出し区間で算出したケフレンシーの平均値において、切り出し区間の最小時間におけるケフレンシー値が最大となるケフレンシー分布※

$$x(n), (0 \leq n \leq N)$$

【0051】

★ ★ 【数2】

$$W_H(n) = 0.54 - 0.46 \cos\left(\frac{2n\pi}{N-1}\right)$$

※パターン(特徴量9)、及び時間成分毎のケフレンシー平均値の分布パターン(特徴量10)、場面先頭200msec分の音量平均(特徴量11)、場面末尾200msec分の音量平均(特徴量12)、場面先頭500msecから1000msecまでの音量平均(特徴量13)、場面末尾1000msec前から500msec前までの音量平均(特徴量14)、7.5kHz以上のサブバンド音に対しては、さらに細かい周波数に分割し、各サブバンドの高域音量が0以上となる継続時間(特徴量15)を抽出する。

【0048】なお、特徴量6において、可聴音の帯域として、20kHzまでの帯域を用いたが、20kHzを超えた帯域を用いても同様に実施可能である。また、32のサブバンドに分割しているが、分割するサブバンド数を任意の値にしても同様に実施可能である。特徴量11から特徴量14までに示した音量の切り出し時間に任意の値を用いても同様に実施可能である。

【0049】ここで、ケフレンシーの算出方法について、図3のフローチャートを用いて説明する。音の標本値系列を切り出す(ただし、標本化された時系列を $x(n)$, $(0 \leq n \leq N)$ と表す:ここで N は標本数を表す)(ステップ1)。切り出した区間の両端に急激な変化が起こらないように(数2)で示すハミング窓関数 $W_H(n)$ を乗じる(ステップ2)。

【0050】

【数1】

$$\dots(1)$$

$$\dots(2)$$

【0052】高速フーリエ変換アルゴリズムを用いて、(数3)に示す離散的フーリエ変換 $S(\omega)$ を計算する(ステップ3)。この離散的フーリエ変換を施された値 $S(\omega)$ の絶対値をとり $(|S(\omega)|)$ (ステップ4)、対数変換を行なう $(\log|S(\omega)|)$ (ステップ5)。(数*

$$S(\omega) = \frac{1}{2\pi N} \left| \sum_{n=0}^{N-1} x(n) e^{-j\omega n} \right|^2 \quad \dots(3)$$

【0054】

※ ※ 【数4】

$$C_n = \frac{1}{N} \sum_{n=0}^{N-1} \log |S(\omega)| e^{j2\pi kn/N}, (0 \leq n \leq N-1) \quad \dots(4)$$

【0055】なお、以上の説明で用いた特徴量以外に、従来さまざまな特徴量が提案されており、本発明において、その他の特徴量についても同様に実施可能である。

【0056】次に、特徴量選択部5300の特徴量を選択し、パターン化する処理について説明する。特徴量選択部5300は、特徴抽出処理部5200の分析によって得られた複数の特徴量の内、データ入力部1000から指示のあった特徴量を選択し特徴量重み学習部5400に出力する。

【0057】特徴量選択部5300の特徴量の選択は、データ入力部1000の指示により行われる。例えば、全ての特徴量を用いる指示や、特徴量1だけを用いる指示などがある。選択された特徴量をパターン化する処理を特徴量1から特徴量15を用いて説明する。

【0058】特徴量1として抽出される16色ヒストグラムに対し、頻度の最大値が1、最小値が0となるように規格化を行なった後、16要素の入力パターンを生成する。さらに、この16色ヒストグラムの頻度分散を計算し、1要素の入力パターンを生成する。従って、特徴量1からは最大17要素のパターンが生成される。なお、ヒストグラムの色の種類は、色の値の量子化値を変更する等の方法によって、任意の種類に変更でき、ヒストグラムの色の種類に任意の値を用いても同様に実施可能である。

【0059】特徴量2については、フレームを構成する全画素に対するエッジ画素の割合を算出し、全ての場面について、最大となる割合が1、最小となる割合が0となるように規格化した値を1つのパターンとする。また、フレームを縦横3分割、計9分割し、それぞれの分割場面に含まれる画素に対するエッジ画素の割合を算出する。さらに、全ての場面について、それぞれの分割画面のエッジ画素割合の最大値が1、最小値が0となるように規格化したものをパターンとする。従って、特徴量2からは10パターンが生成される。なお、画面の分割数に任意の値を用いても同様に実施可能である。

【0060】特徴量3、4、5は対象となるフレームが異なるだけでパターン化する処理は同様である。フレームを構成する全画素に対する変化のあった画素(エッジ画素がエッジでなくなった画素/エッジでない画素がエ

*4)に示す逆周波数変換した値 C_n がケプストラムである(ステップ6)。

【0053】

【数3】

ッジになった画素)の割合を算出し、全ての場面について、最大となる割合が1、最小となる割合が0となるように規格化した値を1つのパターンとする。また、フレームを縦横3分割、計9分割し、それぞれの分割場面に含まれる画素に対する変化のあった画素の割合を算出する。さらに、全ての場面について、それぞれの分割画面の変化のあった画素割合の最大値が1、最小値が0となるように規格化したものをパターンとする。

【0061】特徴量6については、32のサブバンド毎に音量平均の最大値が1、最小値が0となるように規格化を行なう。同様に、特徴量7は音量最大値について、特徴量8は、音量分散について規格化を行なう。

【0062】特徴量9については、ケフレンシー成分毎にそれぞれの最大値が1、最小値が0となるように規格化する。ただし、ケフレンシー成分が16を超える場合には、隣接する成分の平均値等を用いて16を超えないようにし、規格化を行なう。

【0063】特徴量10についても同様の動作を行なう。なお、ケフレンシー成分のしきい値として16以外の値を用いても同様に実施可能である。

【0064】特徴量11については、場面先頭200msec分の音量平均の最大値が1、最小値が0となるように規格化する。特徴量12、13、14についても対象となる音量が異なるだけで、同様の動作を行なう。

【0065】特徴量15についても、高域音持続時間の最大値が1、最小値が0となるように規格化する。

【0066】なお、以上の説明で用いた特徴量以外の特徴量についても同様に実施可能である。

【0067】次に、特徴量重み学習部5400、自動分類部5500及び分類結果評価部5600の動作について、図4のフローチャートを用いて説明する。

【0068】上記合致検索キー候補の指定により得られたどこに分類するかを指示した教師データをデータ記憶管理部4000より入力し(ステップ1)、特徴量抽出部5200によって抽出された合致キー候補映像を含む全ての検索キー候補映像の特徴量データを特徴量重み学習部5400に入力する(ステップ2)。

【0069】

【数5】

相関係数

$$c_coef = 1.0 + \left(\left(\frac{P_num}{T_num} - 1.0 \right) \times \alpha \right) \quad \dots(5)$$

【0070】特徴量毎に相関係数 c_coef を(数5)で計算する。ここで、 P_num は、相関のあったデータ数であり、 T_num は相関係数を算出するためのデータ数である。相関係数を算出するためのデータ数は、合致検索キー候補を2つ選ぶ組み合わせの数を $N1$ とし、合致検索キー候補の1つと合致検索キー候補以外の検索キー候補を1つ選ぶ組み合わせ数を $N2$ としたとき、最大 $N1$ と $N2$ の積($N1 \times N2$)になる。相関があるかどうかの判定は、合致検索キー候補の2つの組み合わせにおける差分二乗値をA、合致検索キー候補の1つと合致検索キー候補以外の合致検索キー候補の組み合わせにおける差分二乗値をBとした時 $B-A > 0$ となる組み合わせ数であり、特徴量毎に計算する。 α は学習係数であり、 $0 < \alpha < 1$ を満たすように定める(ステップ3)。

【0071】終了条件1は、特徴量重み学習の繰り返し回数であり、この条件によってあらかじめ設定された回数の学習をおこなう(ステップ4)。既に学習済みの重みを用いて学習するか、初期化するかかの判定を行い(ステップ5)、初期化する場合は、特徴量重みの初期化を行う(ステップ6)。初期化を行わない場合は、既に学習済みの特徴量重みを用いて学習を行う(ステップ7)。

【0072】特徴量重みは学習を重ねることによって、変化していき、ある特徴量の特徴量重みが上限に達したかどうかの判定を行う(ステップ8)。ある特徴量の特徴量重みが上限に達したときには、全ての特徴量重みと学習係数を小さくする。例えば、特徴量重みを2分の1にし、学習係数を10分の1にする(ステップ9)。ステップ4からステップ9までを繰り返し行い、特徴量重みを更新していく。

【0073】ステップ4において設定した繰り返し回数を終了した場合、自動分類部5500では、まず、自己組織化特徴マップ(以下SOM)と呼ばれるニューラルネットワークのパラメータ設定、初期化が行なわれ、その後、設定された繰り返し回数に従い、重みづけされた特徴量(入力ベクトル)を、SOMに繰り返し入力する。入力ベクトルの繰り返し毎に学習は進行し、結果的に、入力ベクトル空間で近い(類似した)パターンは、近い出力ユニットに写像されるようになる。このようにして、入力ベクトルは出力ユニット単位に分類される(ステップ10)。

【0074】この分類結果は、分類結果評価部5600に伝えられ、合致検索キー候補を含むカテゴリにおいて、合致検索キー候補が含まれる割合、あるいは、全ての合致*

$$d1(t) = ((a1 - b1)^2 + (a2 - b2)^2) \quad \dots(6)$$

【0081】合致検索キー候補の中から、1つ、合致検

*検索キー候補に対する同じカテゴリに含まれる合致検索キー候補の割合を計算することによって評価が行われる(ステップ11)。ステップ11の評価結果を用いて終了条件2の判定を行い(ステップ12)、終了条件2を満足していない場合、特徴量重み学習の繰り返し回数(終了条件1)を変更し(ステップ13)、ステップ4に戻る。終了条件2を満足している場合、分類結果を検索キー抽出部5700に出力する。

【0075】ステップ4からステップ13までの動作を繰り返し行うことにより、検索意図と相関のある特徴量の重みが増す結果として、指定された複数の合致検索キー候補を含むカテゴリが同一のカテゴリに近づいていくことになる。ここで、繰り返しの終了条件は、繰り返し回数あるいは分類結果の評価値のしきい値により設定される。

【0076】次に、特徴量重み学習部5400の特徴量重み学習動作について詳細に説明する。特徴量重み学習部5400の特徴量重み学習は、合致検索キー候補を含むカテゴリを検索意図を反映した同一のカテゴリに近づけるために行なう動作である。

【0077】合致検索キー候補を含むカテゴリに含まれる全ての検索キー候補を合致検索キー候補とそれ以外の合致検索キー候補に分け、合致検索キー候補の類似度が大きくなるように、かつ合致検索キー候補とそれ以外の合致検索キー候補の類似度が小さくなるように特徴量重みを更新するものである。

【0078】特徴量重み更新処理を図5のフローチャートを用いて説明する。(ただし、 t 時刻における特徴量重みベクトルを $W(t)$ とし、学習率係数 β を $0 < \beta < 1$ を満たすように定める。)

初期の時刻を $t=0$ とし、特徴量重みベクトル $W(t)$ の初期化を行なう(ステップ1)。終了条件を満たせば動作を停止し、そうでなければステップ3に進む(ステップ2)。合致検索キー候補の中から、ランダムに2つの場面を選択する(ステップ3)。ただし、ステップ3では、2つの場面の選び方として、全ての組合せで選んでも良い。

【0079】時刻 t におけるステップ2で選択された場面の差分の2乗値をそれぞれの特徴量に対し計算する。こうして時刻 t における差分2乗ベクトル $d1(t)$ を(数6)により得られる。(ステップ4)。

【0080】

【数6】

索キー候補以外の合致検索キー候補の中から1つランダ

ムに場面を選択する(ステップ5)。ただし、ステップ5では、合致検索キー候補以外の選び方として均等に選んでも良い。時刻 t におけるステップ4で選択された場面の差分2乗ベクトル $d_2(t)$ を計算する(ステップ*

特徴量重み学習式

$$W(t+1) = \left(1 - \frac{d_j(t) - d_e(t)}{d_j(t) + d_e(t)} \times \beta \right) \times W(t) \times c_coef \quad \dots(7)$$

【0083】特徴量重みベクトルを $W(t+1)$ を(数7)とおく(ステップ7)。次に、 $t = t + 1$ として、ステップ2に戻る。

【0084】このようにステップ2からステップ7までの動作を繰り返し行なうことにより、合致検索キー候補の特徴量重みが大きくなるように、かつ合致検索キー候補以外の合致検索キー候補の特徴量重みが小さくなるように学習されていく。

【0085】なお、終了条件は、次の3つから選択する。(1)特徴量重み学習部5400にあらかじめ設定した、動作の繰り返し回数に達した時、(2)自動分類部5500の分類結果を表示部に表示し、利用者の判断で、データ入力部1000の指示により強制的に終了指示した時、または(3)自動分類部5500で分類された場面のグループに合致検索キー候補が含まれる割合あるいは数が一定値以上に達した時の3つから選択する。

【0086】次に、自動分類部5500の自動分類動作について詳細に説明する。自動分類部5500の自動分類動作の代表例として、SOM法による教師なし競合学習の動作を説明する。SOM法は、ニューラルネットワークの一種であり、1990年にT.Kohonenによって提案された。このSOM法の特徴は、教師なしの学習によって、入力信号パターンの似たもの同士は競合層上の空間的に近い位置に、似ていないもの同士は、遠い位置に写像するような入力信号パターンの組織化を行なうものである。

【0087】まず、SOM法のネットワーク構成図を図6に示し、以下に説明する。ネットワークは、入力層と*

*6)。

【0082】

【数7】

※競合層(出力層)の2層で構成され、入力層を構成するユニット(入力ユニット)は、1つの特徴量に対応し、入力信号ベクトルと同じ数だけ用意され、競合層(出力層)のユニットと全数結合している。一方、競合層(出力層)を構成するユニット(出力ユニット)は、分類されるグループに対応し、2次元格子に規則的に配置される。

【0088】次に、SOM法による学習動作を図7のフローチャートを使って説明する。 t 時刻における入力ユニット j への入力を $X_j(t)$ 、図6(b)に示すように入力ユニット j から出力ユニット k へのリンクの重みを $W_{kj}(t)$ とする。また、図6(a)に示すように出力ユニット k の t 時刻における近傍集合を $N(k, t)$ とし、 $t_1 < t_2$ を満たす任意の時刻 t_1, t_2 に対して、 $N(k, t_1) \subset N(k, t_2)$ が成り立つように時間の経過とともに近傍の領域を小さくしていく。この時、全てのリンクの重み W_{kj} を小さなランダム値に設定する(ステップ1)。

【0089】初期の時刻を $t=0$ とし、全ての出力ユニット k に対して近傍領域 $N(k, 0)$ を定義する。例えば、 k を中心とした大きな六角形を $N(k, t_1)$ とする。ここで、学習率 $\alpha(0)$ を、 $0 < \alpha(0) < 1$ かつ、時間とともに減少していくような関数として定め、例えば(数8)のような関数とする(ステップ2)。

【0090】

【数8】

$$\alpha(t) = \frac{B}{A+t}, (A, B \text{ は定数}, 0 < B < A) \quad \dots(8)$$

【0091】 $t+1$ 番目のデータがなければ動作を終了し、そうでなければ $t+1$ 番目のデータを入力する(ステップ3)。

【0092】出力ユニット k と入力ユニット j との距離 Sim_k を、全ての出力ユニット k に対して、 $X_j(t) - \star$

$$Sim_k = \sum_j (X_j(t) - W_{kj}(t))^2 \quad \dots(9)$$

【0094】

☆ ☆ 【数10】

$$Sim_{k_min} = \min_k (Sim_k) \quad \dots(10)$$

【0095】出力ユニット k と入力ユニット j との最小距離 Sim_{k_min} を、全ての出力ユニット k に対し、ステップ4で計算した値の最小値を(数10)で求め、そ

★ $W_{kj}(t)$ を2乗した和として(数9)で計算する(ステップ4)。

【0093】

【数9】

れにより最小値をもつ出力ユニット k_min が得られる(ステップ5)。

【0096】最終的に、自己組織化のリンク重み $W_{kj}(t)$

+1)として、全ての出力ユニット $k \in N(k_{\min}, t)$ * 【0097】
 について、全ての入力ユニット j に対して(数11)に * 【数11】
 より求める(ステップ6)。 *

$$W_{kj}(t+1) = W_{kj}(t) + \alpha(t)(X_j(t) - W_{kj}(t)) \quad \dots(11)$$

【0098】 $t = t + 1$ とおき、ステップ3に行く(ステップ7)。このようにステップ3からステップ6までの動作を繰り返し行なうことにより、入力ベクトルの分布を代表するような結合重みが、いくつかの出力ユニットに対して形成されるようになる。

【0099】次に、検索キー抽出部5700は、自動分類部5500から出力された学習結果から、合致検索キー候補を含むカテゴリからそのカテゴリを代表する特徴量と重み並びに、自動分類部5500の自己組織化のリンク重みが検索キーとして抽出される。ここで、カテゴリを代表する特徴量とは、合致検索キー候補から抽出した特徴量あるいは合致検索キー候補から抽出した特徴量が最も強く反応した出力ユニットにおいて、合致検索キー候補よりも強く反応した合致検索キー候補の特徴量のことをいう。

【0100】検索キー管理・蓄積部5800は、特徴量重み学習部5400からの合致検索キー候補を含むカテゴリからそのカテゴリを代表する特徴量と重み並び自動分類部5500からの自己組織化のリンク重みが検索キーとして抽出された検索キーを蓄積する。

【0101】なお、合致検索キー候補を含むカテゴリに属する合致検索キー候補を表示部3000に出力した後、データ入力部1000からの指示によっても抽出できる。このようにして抽出された検索キーは、検索キー管理・蓄積部5800に蓄積される。この時、検索キーの再利用に備えて、利用者はデータ入力部1000を用いて、検索キーにタ

グを付けることもできる。

【0102】なお、自動分類部5500を自己組織化マップを用いて動作させる例で説明したが、その他の神経回路網モデルについても同様に実施可能である。

【0103】次に、検索用データ生成部6000について説明する。検索データ生成部6000は、図1に示すように検索対象映像の色やエッジ強度の変化を分析することによって場面の変わり目を検出する場面切替え検出部5100と、場面切替え検出部5100によって検出された場面単位に、場面的特徴量を抽出する特徴量抽出部5200とで構成される。検索用データは、一般的に検索対象となるデータ量を削減するために、検索対象となる映像から特徴的なデータを抽出したものである。

【0104】場面切替え検出部5100は、検索対象となる映像はデータ記憶管理部4000から読み出され、場面切替え検出部5100で検索キー候補映像の場面切替え検出動作を行なうもので、場面切替え検出結果を場面的先頭フレーム番号あるいは場面的先頭の時間として特徴量抽出部5200に出力するものである。特徴抽出処理部5200は、場面単位に映像や音の情報を分析処理し、特徴量選択部53

00に出力するものである。特徴抽出処理部5200の特徴量抽出動作は、場面切替えの単位に、映像、音に関して表1に示すような特徴量が検索用データとして抽出され、再び、データ記憶管理部4000に記憶される。

10 【0105】場面切替え検出部5100と特徴抽出処理部5200は、前述の検索キー生成部5000と同一のものであり、詳細な説明は省略するものとする。

【0106】次に、検索部7000について説明する。検索部7000は、図1に示すように検索キー生成部5000で抽出された検索キーを蓄積・管理する検索キー管理・蓄積部5800と、データ記憶管理部4000から読み出した検索用データを検索キーの重み付けされた特徴量を用いて神経回路網モデルの1つである自己組織化マップと呼ばれるニューラルネットワークを用いて分類する自動分類部5500から構成される。

【0107】検索部7000の検索動作は、予め検索キー生成部5000で生成された検索キーを用いて、入力部1000から検索キーに付けられた場面を反映したタグにより検索指示が入力されることによって開始する。主制御部2000は、検索キーのリンク重みを自動分類部5500に設定し、検索キーの重みづけを行なって、データ記憶管理部4000から読み出した検索用データを検索キーの特徴量とともに自動分類部5500に入力する。自動分類部5500は、学習が終了したニューラルネットワークに、全場面的特徴ベクトルを入力し、それぞれの場面がどの出力ユニットの近傍で最も強く反応するかを検出し、これを分類結果として出力し、表示部3000に表示される。この時検索キーに類似した検索用データは、検索キーの近隣に表示される。この動作を全ての検索用データに対して行なう。

【0108】なお、検出結果に対し、検索意図に合致しているかどうかの評価を行い、検索キー候補映像に対して行なった検索キー生成動作と同様の動作も実施可能である。

【0109】なお、自動分類部に神経回路網モデルを用いた例で説明したが、これに限定されるものではなく、パターン認識等で用いられる各種クラスタリング方法を用いても良い。

【0110】

【発明の効果】以上のように本発明の類似検索システムは、検索意図を特徴づける特徴量の重み学習と、神経回路網モデルを用いた分類及び分類結果の評価を繰り返すことにより、サンプル画像が検索意図を反映した同一のカテゴリとなるように学習で得られた結果を検索キーとして抽出し、検索効率を向上できるという効果が得られる。

【0111】また、検索キーとして、特徴量重みおよび入出力ユニットのリンク重みは学習した結果が検索意図を反映したものであり、学習した結果を検索キーとして利用できる。

【図面の簡単な説明】

【図1】本発明の実施の形態の類似検索システムの構成図

【図2】同場面検出部の類似検索システムの場面切替え検出動作を示すフローチャート

【図3】同類似検索システムの特徴量抽出部のケプレンシーの算出動作を示すフローチャート

【図4】同類似検索システムの特徴量重み学習部、自動分類部及び分類結果評価部の動作を示すフローチャート

【図5】同類似検索システムの特徴量重み学習部の特徴量重み更新処理のフローチャート

【図6】同類似検索システムの自動分類部のSOM法のネットワーク構成図

【図7】同類似検索システムの自動分類部のSOM法の学習動作のフローチャート

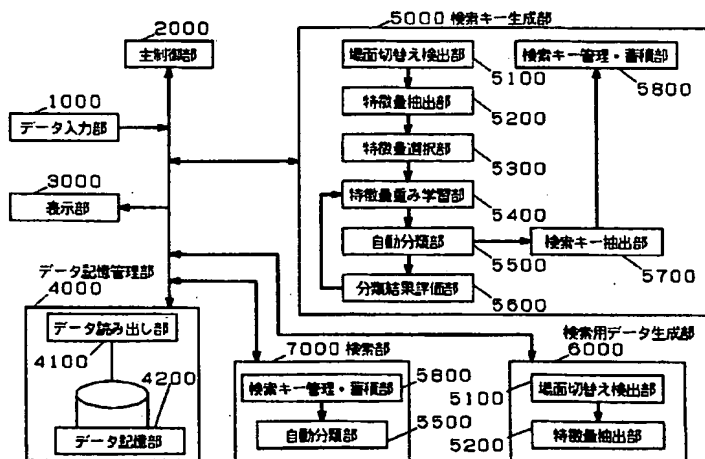
【図8】従来の類似検索システムのブロック構成図

【図9】従来の類似検索システムの動作説明のフローチャート

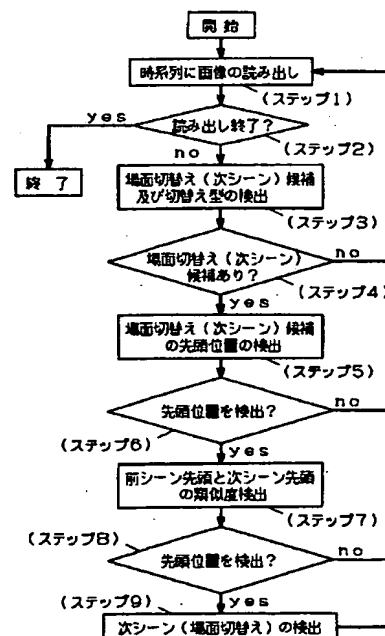
【符号の説明】

1000	データ入力部
2000	主制御部
3000	表示部
4000	データ記憶管理部
4100	データ読み出し部
4200	データ記憶部
5000	検索キー生成部
5100	場面切替え検出部
5200	特徴量抽出部
5300	特徴量選択部
5400	特徴量重み学習部
5500	自動分類部
5600	分類結果評価部
5700	検索キー抽出部
5800	検索キー管理・蓄積部
6000	検索用データ生成部
7000	検索部

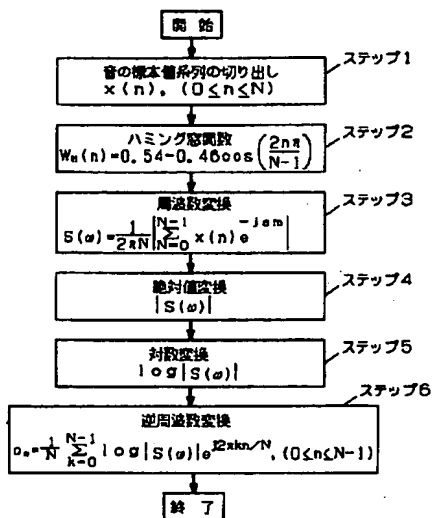
【図1】



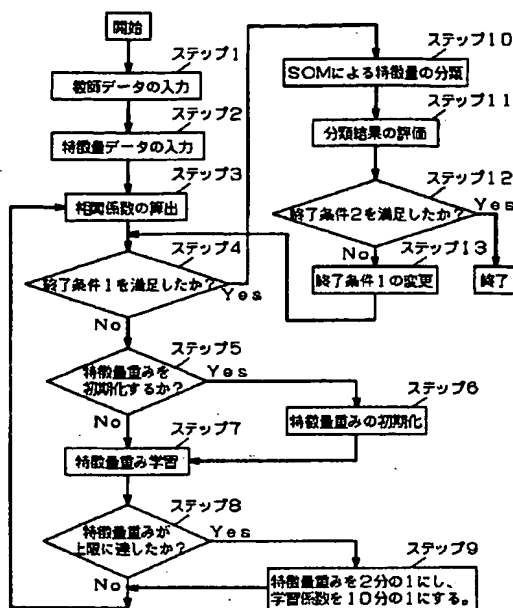
【図2】



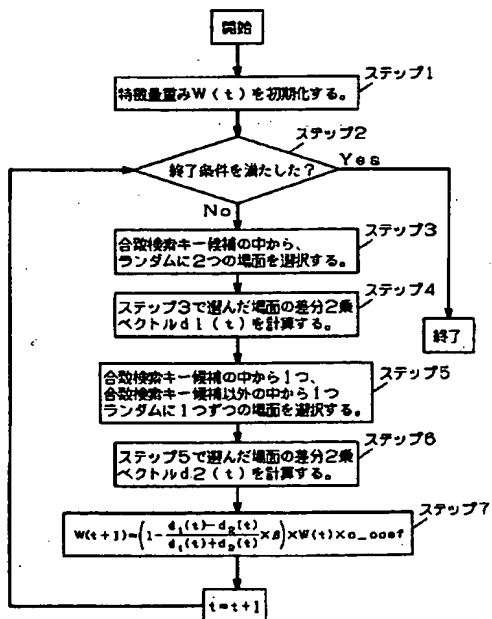
【図3】



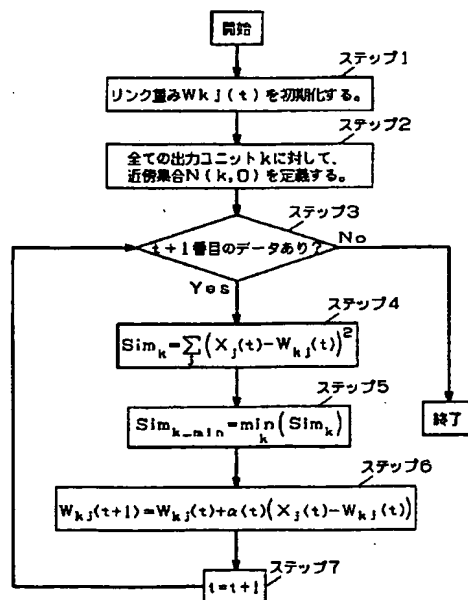
【図4】



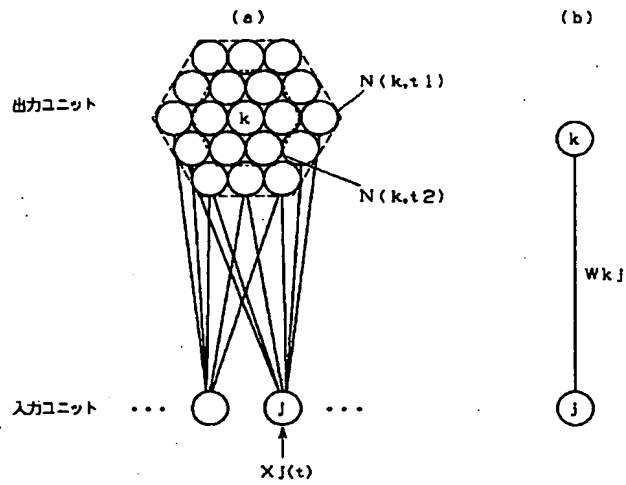
【図5】



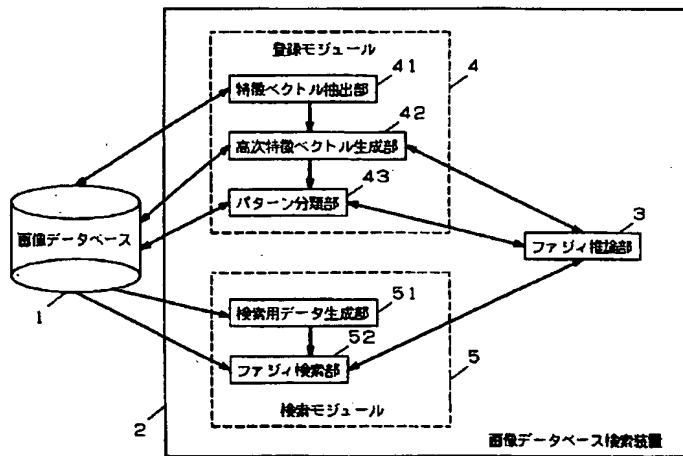
【図7】



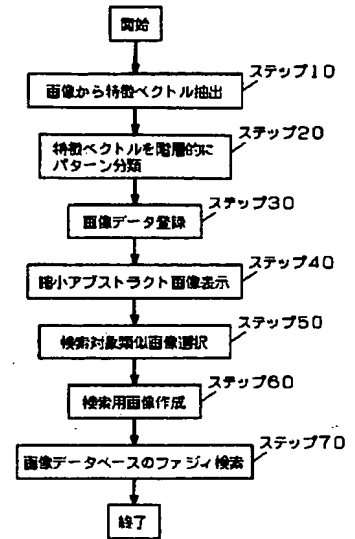
【図6】



【図8】



【図9】



フロントページの続き

(51) Int. Cl. 6

識別記号

F I

G 0 6 F 15/403

3 5 0 C

15/70

4 6 5 A